Attorney Docket Number **IN01164K**

PATENT APPLICATION

ISOLATION OF *MICROMONOSPORA CARBONACEA* VAR *AFRICANA* pMLP1 INTEGRASE AND USE OF INTEGRATING FUNCTION FOR SITE-SPECIFIC INTEGRATION INTO *MICROMONOSPORA HALOPHITICA* AND *MICROMONOSPORA CARBONACEA* CHROMOSOME

INVENTORS:

Thomas J. Hosted, Jr., a citizen of the United States, residing at 160 Summit Avenue, Apt. 302, Summit, NJ 07901, U.S.A.

Ann C. Horan, a citizen of the United States, residing at 144 Mountain Avenue, Summit, NJ 07901, U.S.A.

ASSIGNEE: Schering Corporation

"Express Mail" Label No. EL226886013US
Date of Deposit: May 15, 2001

IMMAC J. THAMPOE, Reg. No. 36322
Schering-Plough Corporation
Patent Department, K-6-1, 1990
2000 Galloping Hill Road
Kenilworth, New Jersey 07033-0530
Telephone No.: (908) 298-5061
Facsimile No.: (908) 298-5388

ISOLATION OF *MICROMONOSPORA CARBONACEA* VAR AFRICANA pMLP1 INTEGRASE AND USE OF INTEGRATING FUNCTION FOR SITE-SPECIFIC INTEGRATION INTO *MICROMONOSPORA HALOPHITICA* AND *MICROMONOSPORA CARBONACEA* CHROMOSOME

This application claims the benefit of United States Provisional Application No. 60/204,670 filed May 17, 2000.

Field of the Invention

The present invention relates generally to isolated nucleic acids and the creation of vectors for the study and expression of genes in actinomycetes. The invention more particularly relates to genes isolated from a *Micromonospora* lysogenic phage which can be used to create vectors for site-specific integration into *Micromonospora* chromosomes.

Background

Actinomycetes are branched filamentous Gram-positive bacteria. *Streptomyces, Micromonospora, Nocardia, Actinoplanes, Saccharopolyspora, Actinomadura, Thermomonospora, Microbispora, Streptosporangium* and others all represent genera of the Actinomycetes (Atlas of Actinomycetes, Asakura Publishing Co., Ltd 1996). Actinomycetes are very important industrially because they produce a variety of secondary metabolites such as antibiotics, herbicides, anticancer agents, antihelmintics, and anabolic agents (Demain., Appl. Microbiol and Biotechnology., 1999, **52**:455-463). Antibiotics are a large and complex group of chemical substances which exert deleterious effects on other organisms, many of which organisms are harmful to humans. Thus, antibiotics are particularly important secondary metabolites to study and produce. This is especially true because many pathogens can develop antibiotic resistance to known antibiotics.

Given the actinomycetes' proclivity for producing secondary metabolites such as antibiotics, it is especially advantageous to develop new tools such as vectors, promoters and the like to allow actinomycetes to be easily genetically manipulated. These tools would make it possible to control the levels of expression of genes encoding for secondary metabolites and also would make it possible to prepare derivatives or intermediates of these metabolites. In addition, the development of new vectors utilizing novel genes would make it possible to program microorganisms such as actinomycetes to produce recombinant products such as hybrid antibiotics via genetic engineering techniques.

Integrating vectors are vectors which integrate into a transformed host's chromosome rather than replicating autonomously. They are particularly useful in transforming actinomycetes because they allow for the especially efficient production of secondary metabolites because of their high transformation rates, site specific integrative capacity and stable maintenance in host chromosomes without antibiotic selection.

Vectors have been developed for use in actinomycetes that contain *att/int* functions for site-specific integration of plasmid DNA. The two systems available make use of the *att/int* functions of bacteriophage phiC31 (U.S. Patent No. 5,190,870) and plasmid pSAM2 (U.S. Patent No. 5,741, 675). However, there is a need for additional vectors with *att/int* functions for site-specific integration in *M. carbonacea* and similar organisms.

The present inventors have responded to the above needs and have isolated genes from the actinomycete, *Micromonospora carbonacea var. africana* (ATCC39149, SCC1413) lysogenic phage pMLP1, in order to create vectors which can be used for site-specific integration into *Micromonospora* chromosomes. These integrating vectors can be used to express actinomycete genes, manipulate secondary metabolic pathways and create new metabolic products such as hybrid antibiotics.

Summary of the Invention

The present invention provides novel polynucleotide sequences coding for integrase (*int*) and excisionase (*xis*) genes and an integrase attachment site (*att*P) isolated from pMLP1, a lysogenic phage isolated from *Micromonospora carbonacea var. africana* (ATCC39149, SCC1413). The invention also provides recombinant vectors comprising these genes as well as hosts transformed with these vectors and methods of transforming these hosts.

In one embodiment, the present invention provides isolated polynucleotides comprising sequences which are at least about ninety percent homologous to the nucleotide sequences set forth in SEQ ID NOS: 1-3. These isolated polynucleotides encode novel genes and DNA sequences involved in plasmid integration into a host chromosome. Specifically, these isolated sequences encode a site-specific integrase (*int*), an excisionase (*xis*), and an integrase attachment site (*att*P). In a preferred embodiment, the polynucleotides comprise sequences set forth in SEQ ID NOS: 1-3.

In addition, the invention provides isolated polynucleotides having a sequence at least about 90% homologous to SEQ ID NOS: 4-9, and preferably, having the nucleotide sequences set forth in SEQ ID NOS: 4-9. These sequences encode the *M. carbonacea (att*B) region (SEQ ID NO: 4) as well as the left and right juncture regions *att*B/*att*P (SEQ ID NO: 5) and *att*P/*att*B (SEQ ID NO: 6) regions formed when the *att*P site of pMLP1 is integrated into the *attB* site of *M. carbonacea*. In addition, these sequences encode the *M. halophitica* *att*B region (SEQ ID NO: 7), as well as the left and right juncture regions *att*B/*att*P (SEQ ID NO: 8) and *att*P/*att*B (SEQ ID NO: 9) formed when the *att*P site of pMLP1 is integrated into the (*att*B*)* site of *M. halophitica.*

In another embodiment, the present invention provides recombinant vectors comprising one or more nucleotide sequences which are at least about ninety percent homologous to the nucelotide sequences set forth in SEQ ID NOS: 1-3. In a preferred embodiment, the invention provides a recombinant vector comprising one or more of SEQ ID NOS: 1-3. In an especially preferred embodiment, that vector is an integrating vector capable of integrating into the chromosome of a host cell.

In yet another embodiment, the present invention provides host cells comprising the vectors of the instant invention. In a preferred embodiment, the host cell is bacterial. In an especially preferred embodiment, the host cell is an actinomycete such as a *Micromonospora*.

In a final embodiment, the invention provides a method for transforming an actinomycete with an integrating vector comprising a) isolating a polynucleotide having a sequence at least ninety percent homologous to a sequence selected from the group consisting of SEQ ID NOS: 1-3; b) inserting the polynucleotide or polynucleotides into a vector; and c) transforming an actinomycete such that the vector integrates into the actinomycete chromosome. Preferably the isolated polynucleotides have sequences selected from the group consisting of SEQ ID NOS: 1-3. Most preferably, the vector comprises SEQ ID NOS: 1-3.

Brief Description of the Drawings

Fig. 1 is a schematic of plasmid pSPRH826b, an *E. coli-Micromonospora* insertion vector. β-lactamase, ampicillin resistance; hygromycin resistance; oriT, RK2 origin of transfer for conjugation. Restriction sites are indicated.

Fig. 2 is a schematic of plasmid pSPRH840, which was constructed by inserting the *xis*, *int* and *att*P regions from pMLP1 into backbone pSPRH826b. β-lactamase, ampicillin resistance; hygromycin resistance; oriT, RK2 origin of transfer for conjugation; xis, excisionase; int, integrate; *att*P: attachment site.

Fig. 3 is a schematic representation of pSPRH480 integration into *att*B (tRNA-His) located at the 3' end of the tRNA-His gene. xis, excisionas int, integrase; *att*P, attachment site pSPRH840 indicated by crosshatches; *att*B attachment site chromosome indicated by a solid black line; tRNA-His, transfer RNA histidine; *att*B/*att*P, left integration juncture; *att*P/*att*B, right integration juncture.

Fig. 4 are sequences of pMLP1 *att*P (region containing the *att*P site), *M. carbonacea* *att*B region, pSPRH840 *att*B/*att*P left juncture region, pSPRH840 *att*P/*att*B right juncture region, (labeled a-d, respectively). Regions of *att*P, *att*B, *att*B/*att*P sharing homology are indicated by bold larger sized font. 4a, pSPRH840 *att*P site indicated by large bold font; 4b, *M. carbonacea* *att*B site indicated by large bold font. Arrows indicate: a, *att*P (pSPRH840 attachment site); b, *M. carbonacea* tRNA-His gene and *att*B site; c, pSPRH840 *att*B/*att*P tRNA-His gene; d, pSPRH840 *att*P/*att*B 3' region of tRNA-His gene. Inverted repeats are indicated by small arrows.

Fig. 5 are sequences of pMLP1 *att*P (region containing the *att*P site), *M. halophitica* *att*B region, pSPRH840 *att*B/*att*P left juncture region, pSPRH840 *att*P/*att*B right juncture region (labeled a-d, respectively). Regions of *att*P, *att*B, *att*B/*att*P and *att*P/*att*B sharing homology are indicated by bold larger sized font. 5a, pSPRH840 *att*P site indicated by large bold font; 5b, *M. halophitica* *att*B site indicated by large bold font. Arrows indicated: a, *att*P (pSPRH840 attachment site); b, *M. halophitica* tRNA-His gene and *att*B site; c, pSPRH840 *att*B/*att*P tRNA-His gene; d, pSPRH840 *att*P/*att*B 3' region of tRNA-His gene. k Inverted repeats are indicated by small arrows.

Detailed Description of the Invention

The present invention relates to nucleic acids isolated from bacteriophage (lysogenic phage) pMLP1 isolated from *Micromonospora carbonacea* var. *africana* ATCC39149. In addition, the invention relates to vector constructs made utilizing these nucleic acids. Specifically, these vector constructs can be utilized to integrate in a site-specific manner into the host chromosome of an actinomycete.

Before describing the invention in detail, the following definitions are provided to aid in an understanding of the specification and claims:

"Nucleic acid" or "polypeptide" as used herein refers to purine- and pyrimidine-containing or amino acid polymers of any length, either polyribonucleotides or polydeoxyribonucleotides or mixed polyribo-polydeoxyribo nucleotides. This includes single- and double-stranded molecules, *i.e.*, DNA-DNA, DNA-RNA and RNA-RNA hybrids, as well as "protein nucleic acids" (PNA) formed by conjugating bases to an amino acid backbone. This also includes nucleic acids containing modified bases.

An "open reading frame" (ORF) as used herein is a region of a polynucleotide sequence that encodes a polypeptide; this region may represent a portion of a coding sequence or comprise a total coding sequence for the polypeptide.

A "coding sequence" or a "protein-coding sequence" is a polynucleotide sequence capable of being transcribed into mRNA and/or capable of being translated into a polypeptide. The boundaries of the coding sequence are typically determined by a translation start codon at the 5'-terminus and a translation stop codon at the 3'-terminus.

A "complement" of a nucleic acid sequence as used herein refers to the "antisense" sequence that participates in Watson-Crick base-pairing with the original sequence.

An "isolated" nucleic acid or polypeptide as used herein refers to a nucleic acid that is removed from its original environment such as, for example, from *Micromonosporacea carbonacea* plasmid pMLP1.

A nucleic acid or polypeptide sequence that is "derived from" a designated sequence refers to a sequence that corresponds to a region of the designated sequence. For nucleic acid sequences, this encompasses sequences that are homologous or complementary to the sequence, as well as "sequence-conservative variants" and "function-conservative variants." For polypeptide sequences, this encompasses "function-conservative variants." Sequence-conservative variants are those in which a change of one or more nucleotides in a given codon position results in no alteration in the amino acid encoded at that position. Function-conservative variants are those in which a given amino acid residue in a polypeptide has been changed without altering the overall configuration and function of the native polypeptide, including, but not limited to, replacement of an amino acid with one having similar physico-chemical properties (such as, for example, acidic, basic, hydrophobic, and

the like). "Function-conservative" variants also include any polypeptides that have the ability to elicit antibodies specific to a designated polypeptide.

A "*M. carbonacea*-derived" nucleic acid or polypeptide sequence may or may not be present in other bacterial species, and may or may not be present in all *M. carbonacea* strains. This term is intended to refer to the source from which the sequence was originally isolated. An *M. carbonacea* plasmid-derived polypeptide, as used herein, may be used to search for homologous proteins in other species of bacteria or in eukaryotic organisms such as fungi and humans, etc.

A "probe" refers to a nucleic acid or oligonucleotide that forms a hybrid structure with a sequence in a target region due to complementarity of at least one sequence in the probe with a sequence in the target protein.

Nucleic acids are "hybridizable" to each other when at least one strand of nucleic acid can anneal to another nucleic acid strand under defined stringency conditions. Stringency of hybridization is determined, by the temperature at which hybridization and/or washing is performed and the ionic strength and polarity of the hybridization and washing solutions, as well as other parameters. Hybridization requires that the two nucleic acids contain substantially complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementarity, variables well known in the art.

"Gene" refers to a DNA sequence that encodes through its template or messenger RNA a sequence of amino acids characteristic of a specific peptide, polypeptide or protein. The term "gene" as used herein with reference to genomic DNA includes intervening, non-coding regions, as well as regulatory regions, and can include 5' and 3' ends.

"Gene sequence" refers to a DNA molecule, including both a DNA molecule which contains a non-transcribed or non-translated sequence. The term is also intended to include any combination of gene(s), gene fragment(s), non-transcribed sequence(s) or non-translated sequence(s) which are present on the same DNA molecule.

"Homologous nucleic acid sequences" are those which when aligned and compared exhibit significant similarities. Standards for homology in nucleic acids are either measures for homology generally used in the art by sequence comparison or based upon hybridization conditions, which are described in greater detail below. Nucleotide sequence homology is

observed when there is identity in nucleotide residues in two sequences (or in their complementary strands) when optimally aligned to account for nucleotide insertions or deletions. Substantial homology also exists when one sequence will hybridize under selective hybridization conditions to another. Typically, selective hybridization will occur when there is at least about 55% homology over a stretch of at least about 30 nucleotides, preferably at least about 65% over a stretch of at least about 25 nucleotides, more preferably at least about 75%, and most preferably at least about 90% over about 20 nucleotides. See, e.g., Kanehisa, *Nucleic Acids Res. 12*:203 (1984).

The nucleotide sequences of the present invention may be derived from a variety of sources including DNA, cDNA, synthetic DNA, synthetic RNA or combinations thereof. Such sequences may comprise genomic DNA which may or may not include naturally occurring introns. Moreover, such genomic DNA may be obtained in association with promoter regions or poly (A) sequences. The sequences, genomic DNA or cDNA may be obtained in any of several ways. Genomic DNA can be extracted and purified from suitable cells by means well known in the art. Alternatively, mRNA can be isolated from a cell and used to produce cDNA by reverse transcription or other means.

"cDNA" refers to complementary or copy DNA produced from an RNA template by the action of RNA-dependent DNA polymerase (reverse transcriptase). Thus, a "cDNA clone" means a duplex DNA sequence complementary to an RNA molecule of interest, carried in a cloning vector or PCR amplified. This term includes genes from which the intervening sequences have been removed.

"Recombinant DNA" means a molecule that has been recombined by *in vitro* splicing of cDNA or a genomic DNA sequence.

"Cloning" refers to the use of *in vitro* recombination techniques to insert a particular gene or other DNA sequence into a vector molecule. In order to successfully clone a desired gene, it is necessary to use methods for generating DNA fragments, for joining the fragments to vector molecules, for introducing the composite DNA molecule into a host cell in which it can replicate, and for selecting the clone having the target gene from amongst the recipient host cells.

"Host" includes prokaryotes and eukaryotes. The term includes an organism or cell that is the recipient of a replicable expression vehicle.

An "integrating vector" is a vector capable of site-specific integration into a bacterial chromosome, and specifically into the attB site.

A "shuttle vector" is a vector capable of replication in *E. coli* and a second bacterial strain such as an actinomycete.

The methods traditionally used in molecular biology, such as preparative extractions of plasmid DNA, centrifugation of plasmid DNA in a cesium chloride gradient, agarose or acrylamide gel electrophoresis, purification of DNA fragments by electroelution, protein extractions with phenol or phenol/chloroform, ethanol or isopropanol precipitation of DNA in a saline medium, transformation in *Escherichia coli*, and the like, are well known to a person skilled in the art and are amply described in the literature. Maniatis T., *et al.*, "Molecular Cloning, a Laboratory Manual," Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982; Ausubel F. M., *et al.*, (eds), "Current Protocols in Molecular Biology," John Wiley & sons, New York 1987.

Protocols have been developed to genetically manipulate actinomycete genomes and biosynthetic pathways. These include the construction of *E. coli*-actinomycete shuttle vectors, gene replacement systems, transformation protocols, transposing mutagenesis, insertional mutagenesis, integration systems and heterologous host expression. These techniques are reviewed in numerous articles (Baltz *et al.*, Trends Microbiol., 1998, **2**:76-83, Hopwood *et al.*, Genetic Manipulation of Streptomyces: A Laboratory Manual, 1985; Wohlleben *et al.*, Acta Microbiol. Immunol. Hung, 1994, **41**:381-9 [Review]).

The development of vectors for the genetic manipulation of actinomycetes began with the observation of plasmids in actinomycetes and the development of a transformation protocol of actinomycete protoplasts using polyethylene glycol (Bibb *et al.*, Nature, 1980, **284**:526-31). Many standard molecular techniques for *Streptomyces* were developed by Hopwood for *Streptomyces coelicolor* and *Streptomyces lividans* (Hopwood *et al.*, Genetic Manipulation of Streptomyces: A Laboratory Manual, 1985). These techniques have been adapted and expanded to other actinomycetes.

Vectors incorporating antibiotic-resistance markers (AmR: apramycin; ThR: thiostrepton; SpR: spectinomycin) that function in *Streptomyces* and other features have allowed the development of vectors for (a) integration via homologous recombination between cloned DNA and *Streptomyces* chromosomes, (b) *E. coli*-actinomycete shuttle vectors, and (c) site-specific integration vectors utilizing *att/int* functions from bacteriophage

phiC31 which integrates into the phiC31 *att*B site (U.S. Patent No. 5,190,870) or *att/int* functions from pSAM2 which integrates into the pSAM2 *att*B site (U.S. Patent No. 5,741,675), and (d) gene replacement vectors. Homologous recombination between the cloned DNA and the chromosome can be used to make insertional knockouts of specific genes. *E. coli*-actinomycete shuttle vectors can be used to introduce copies of genes into actinomycetes. Site-specific integration plasmids can be used to introduce heterologous genes into the actinomycete chromosome for complementation, expression studies and production of hybrid secondary metabolites.

Many actinomycetes contain restriction systems that limit the ability to transform organisms by protoplast transformation. More recent gene transfer procedures have been developed for introducing DNA into *Streptomycetes* by conjugation from *Escherichia coli*. This employs a simple mating procedure for the conjugal transfer of vectors from *E. coli* to *Streptomyces* species that involves plating of the donor strain and either germinated spores or mycelial fragments of the recipient strain. Conjugal plasmids contain the 760-bp oriT fragment from the IncP plasmid, RK2 and are transferred by supplying transfer functions in trans by the *E. coli* donor strain. Other recent developments that increase the frequency of recombination of non-replicating plasmids into the recipient actinomycete chromosome include transformation of non-replicating plasmids into protoplasts using denatured plasmid DNA (Oh and Chater, J. Bacteriol., 1997, **179**:1227) and conjugation of non-replicating plasmids from a methyl minus strain of *E. coli*. (Smith *et al.*, FEMS Microbiol. Lett., 1997, **155**:2239).

Various strategies have been used to obtain gene replacements in *Streptomycetes*, for the construction of mutations and the modification of biosynthetic pathways (Baltz *et al.*, 1998, *supra*; Hopwood *et al.*, supra; Wohllenben *et al.*, 1994, *supra*; Baltz and Hosted, TIBTECH, 1996, **14**:245; Baltz, Curr. Op. Biotech., 1990, **1**:1220). These methods have typically employed a two or three step procedure that results in allelic exchange. Initial crossover events between a non-integrating phage, non-replicating plasmid, or temperature sensitive plasmid and the *Streptomycete* chromosome are selected for by antibiotic resistance. Subsequent recombination events that result in gene replacement can be detected by screening the progeny of the initial recombinants by PCR analysis, Southern analysis, appearance of an expected phenotype or screening for the loss of a resistance marker which had previously been exchanged into the loci to be replaced. The last of these methods has

been employed by Khosla *et al.*, Mol. Microbiol., 1992, **6**:323749; Khosla *et al.*, J. Bacteriol., 1993, **175**:2197204, to successfully modify the polyketide biosynthetic route of S. coelicolor. The strategy employed by Khosla *et al.*, 1992, supra, also has the advantage of allowing placement of non-selectable and phenotypically silent alleles into chosen positions of the chromosome. Donadio *et al.*, Proc. Natl. Acad. Sci. U.S.A., 1993, **90**:711923 has also successfully reprogrammed the erythromycin pathway of *Saccharopolyspora erythrae* by gene replacement.

Non-replicating plasmids for gene replacement were initially utilized by Hilleman *et al.*, Nucleic Acids Res., 1991, **19**:72731, who used a derivative of pDH5 to construct mutations in the phosphinothricin tripeptide biosynthetic pathway of *S. hygroscopicus*. Plasmid-insertion events were obtained by thiostrepton selection, subsequent screening of the primary recombinants indicated that 4 of 100 isolates had undergone a double-crossover gene replacement.

Use of counterselectable or negative selection markers such as rpsL (confers streptomycin sensitivity) or sacB (confers sucrose sensitivity) have been widely employed in other microorganisms for selection of recombination that results in gene replacement. In *S. coelicolor*, Buttner utilized *glk* as a counterselectable marker in *att* minus phiC31 phage to select for recombination events to construct gene replacement mutants of three *S. coelicolor* RNA polymerase sigma factors (Buttner *et al.*, J. Bacteriol., 1990, **172**:336778). Hosted has developed a gene replacement system utilizing the rpsL gene for counterselection (Hosted and Baltz, J. Bacterial, 1997, **179**:1806).

The construction of recombinant actinomycete strains to produce hybrid secondary metabolites has been accomplished (Baltz, Antibiotic Resistance and antibiotic development" Harvard Academic Publishers (in press). Current procedures use recombinant DNA techniques to isolate and manipulate secondary metabolic pathways and to express these pathways in surrogate hosts such as *Streptomyces lividans*. Heterologous expression of diverse pathways, polyketide, oligopeptide and ß-lactam biosynthetic pathways, has been achieved. Furthermore, novel polyketide structures have been generated through the manipulation of polyketide genes forming chimeric pathways. Recently, novel polyketide modules have been isolated from environmental sources using PCR amplification and expressed in *Streptomyces* to yield novel chemical structures (Strohl *et al.*, J. Industr.

Microbiol., 1991, **7**:163; Kim *et al.*, J. Bacteriol., 1995, **77**:1202; Ylihonko *et al.*, Microbiology, 1996, **142**:1965).

A number of Actinomycetes harbor integrative "elements" that contain *att/int* functions capable of directing site-specific recombination into the chromosome. These integrative elements include plasmids and bacteriophages which are often capable of transfer and integration into hosts devoid of the integrative-element. In some cases the integrative-element coexists a both a freely replicating and integrated form. *Att/int* regions characteristically consist of an excisionase (xis), integrase (int), and a short DNA region designated the attP element. The integrase acts as site-specific DNA recombinase that directs strand exchange recombination of the attP element with a chromosomal attB site. Often actinomycete attB sites are tRNA genes that share a short segment of identity with the attP element [Reiter *et al.*, "Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements" *Nucleic Acids Res* **17**(5):1907-14 (1989)] that extends from the anticodon loop to the 3' end of the tRNA gene. Recombination of the attP element with the attB site forms an attP/attB juncture (attL) that regenerates a functional tRNA gene and an attB/attP juncture (attR).

Actinomycetes integrative plasmids include *Streptomyces ambofaciens* ATCC23877 pSAM2 [Pernodet *et al.*, "Plasmids in different strains of Streptomyces ambofaciens: free and integrated form of plasmid pSAM2" *Mol Gen Genet* **198**(1):35-41 (1984)], *Saccharopolyspora erythrea* pSE21, *Amycolytopsis mediteranei* pMEA100 [Moretti *et al.*, "Isolation and characterization of an extrachromosomal element from Nocardia mediterranei" *Plasmid* **14**(2):126-33 (1985)], *S. glaucescens* pIJ408 *S. coelicolor* A3(2)SLP1 [Bibb *et al.*, "Excision of chromosomal DNA sequences from Streptomyces coelicolor forms a novel family of plasmids detectable in Streptomyces lividans" *Mol Gen Genet* **184**(2):230-40 (1981)], *A. methanolitica* pMEA300 [Vrijbloed *et al.*, "A plasmid from the methylotrophic actinomycete Amycolatopsis methanolica capable of site-specific integration" *J Bacteriol* **176**(22):7087-90 (1994)], and others. Actinomycete integrative bacteriophages include phiC-31, a broad host-range temperate streptomycete phage [Lomovskaya *et al.*, "Characterization of temperate actinophage phi C31 isolated from Streptomyces coelicolor A3(2)" *J Virol* **9**(2):258-62 (1972)], the *S. rimosus* RP2 and RP3 temperate phages [Rausch *et al.*, "The temperate phages RP2 and RP3 of Streptomyces rimosus" *J Gen Microbiol* **139**(Pt 10):2517-24 (1993)], and the VWB temperate phage from

S. venezuelae [Van Mellaert *et al.*, "Site-specific integration of bacteriophage VWB genome into Streptomyces venezuelae and construction of a VWB-based integrative vector" *Microbiology* **144**(Pt 12):3351-8 (1998)]. The attB sites have been characterized for pSAM2, pSE211, RP3 and VWB and all correspond to the 3' end of a tRNA gene that shares a 58 bp to 112 bp segment of identity with the corresponding attP element. The conservation of tRNA genes at the sequence level often allows integration of these vectors into phylogeneticaly diverse hosts. This is exemplified by pSAM2 *att/int* function derived vectors which can integrate into numerous *Streptomyces* species [Simonet *et al.*, "Excision and integration of a self-transmissible replicon of Streptomyces ambofaciens" *Gene* **59**(1):137-44 (1987); Kuhstoss *et al.* "Site-specific integration in Streptomyces ambofaciens: localization of integration functions in S. ambofaciens plasmid pSAM2" *J Bacteriol* **171**(1):16-23 (1989); Boccard *et al.*, "Structural analysis of loci involved in pSAM2 site-specific integration in Streptomyces" *Plasmid* **21**(1):59-70 (1989)] and *Mycobacterium smegmatis* [Martin *et al.*, "Site-specific integration of the Streptomyces plasmid pSAM2 in Mycobacterium smegmatis" *Mol Microbiol* **5**(10):2499-502 (1991)].

The instant invention relates to the isolation and identification of novel genes from the *M. carbonacea* pMLP1 bacteriophage. These genes have been used to create vectors for site-specific integration into host chromosomes. Specifically, use of the pMLP1 *att/int* site-specific integration function will allow for increasing a given gene dosage and for adding heterologous genes that lead to the formation of new products such as hybrid antibiotics. This procedure has many advantages over methods involving autonomously replicating plasmids. In particular, replicating plasmids require selection to be maintained and control of plasmid copy number is difficult so that gene dosage cannot be controlled. pMLP1 derived vectors integrate as a single copy per chromosome.

Vectors comprising the site-specific integrating function of pMLP1 can be used to permanently integrate copies of the gene of choice into the chromosome of actinomycetes. Vectors lacking actinomycete origins of replication can only exist in their integrated form in actinomycetes. Integrated vectors are extremely stable which allows the gene copies to be maintained without antibiotic selective pressure. The site-specific nature of the integration allows analysis of the integrants.

I. NUCLEIC ACID SEQUENCES

The present inventors have isolated novel genes from *M. carbonacea* bacteriophage pMLP1. Example 1 describes the construction of a *M. carbonacea* cosmid library and creation of plasmid pSPR150 which contains the *M. carbonacea int*, *xis* and *att*P sites.

EXAMPLE I
Construction of a *M. carbonacea* cosmid library and isolation of pSPR150

M. carbonacea chromosomal DNA was partially digested with *Sau3*A1 to yield DNA of ~40 kb in size, treated with alkaline phosphatase (Boehringer Mannheim Biochemicals), ligated to *Bam*HI digested pSupercos II (Stratagene), and packaged with Gigapack II packaging extract (Stratagene). Packaged DNA was titred on *E. coli* XL1-Blue-MFR' (Stratagene) and individual cosmid clones were stored as an ordered array in 96 well microtitre plates. Primary screening filters were prepared using a 96-well dot blot apparatus. Twelve cultures from a row of microtitre wells were pooled, plasmid DNA was prepared, stored as mixed pools and bound to nylon filters (BioRad Zeta-probe GT). Secondary screening was performed on individual cosmid clones from microtitre wells by PCR or Southern analysis.

Degenerate PCR primers PR144 (5' TGCTTCGACGCCATCARGG3') and PR145 (5'GTGGAAICCGCCGAAKCCGC3') were designed to amplify polyketide synthetase type I genes (Hutchinson *et al.*, *Annual Review of Microbiology,* 1995, **49**:201-238). PCR primers PR144 and PR145 were used to amplify a 0.6 kb fragment from *M. carbonacea* chromosomal DNA. The 0.6 kb fragment was cloned into the pNOTA vector (5 Prime 3 Prime Inc., Boulder, Co) and sequence analysis of the insert revealed BLAST homology to polyketide type I genes. PCR analysis of the *M. carbonacea* cosmid library using PR144 and PR145 were used to isolate cosmid pSPR150. The 4 kb pSPR150 insert was sequenced and revealed numerous phage like genes including genes and DNA regions with homology to database integrases, excisionases and *att*P attachment sites.

After obtaining the complete sequence of a 4 kb fragment subcloned from pSPR150 from the Gene Inspector program (Textco, Inc. West Lebanon, New Hampshire) and BLAST analysis (Genetics Computer Group, Program Manual for the GCG Package, Version 7, Madison, Wisconsin) was used to analyze the sequence. This analysis revealed the *int* gene which showed homology to other integrases in the NRRL database. Specifically, the *int*

gene had a BLAST score of 2.6e-31 to Mycobacterium phage Ms6 integrase (accession number AF030986). Analysis of the predicted *att*P site showed homology to the *att*P sites found in phage phiC31 and plasmid pSAM2. In addition, an excisionase (*xis*) gene was identified via BLAST homologies. Specifically, the *xis* had a BLAST score of 0.51 to the c2 bacteriophage excisionase (accession number X94331). While the genes and DNA regions isolated from pMLP1 share homologies with other streptomycete plasmid functions, phylogenetic analysis of the pMLP1 genes and DNA regions clearly indicate that the identified pMLP1 proteins and DNA regions represent novel *Micromonosporacea*-specific pMLP1 bacteriophage functions. Specifically these include novel genes or DNA regions involved in pMLP1 specific integration and excision (*int*, *xis*, *att*P site).

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optical alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. **2**:482, by the homology alignment algorithm of Needlman and Wunsch (1970) J. Mol. Biol. **48**:443, by the search for similarity method of Pearson and Lipman (1988) Proc. Nat'l Acad. Sci. USA **85**:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally Ausubel *et al.*, supra).

An example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described Altschul, *et al.* (1990) J. Mol. Biol. **215**:403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http:www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul, *et al.*, supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs

containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) Proc. Nat'l Acad. Sci. USA **89**:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) Proc. Nat'l Acad. Sci. USA **90**:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Novel genes and DNA regions from the *M. carbonacea* bacteriophage pMLP1 have been isolated. The sequences of the present invention include the specific nucleic acid sequences set forth in the Sequence Listing that forms a part of the present specification. For convenience, the sequences are designated SEQ ID NO: 1 - SEQ ID NOS: 3. The invention encompasses each sequence individually, as well as any combination thereof.

The gene sequences of this invention (SEQ ID NOS: 1-3) encode for both proteins and non-translated DNA regions involved in pMLP1 integration and excision. These sequences all represent novel pMLP1 bacteriophage functions or regions. In addition, sequences of the invention include the *M. carbonacea (att*B) region (SEQ ID NO: 4) as well as the left and right juncture regions *att*B/*att*P (SEQ ID NO: 5) and *att*P/*att*B (SEQ ID NO: 6) regions formed when the *att*P site of pMLP1 is integrated into the *attB* site of *M. carbonacea*. Also, these sequences encode the *M. halophitica att*B region (SEQ ID NO: 7), as well as the left and right juncture regions *att*B/*att*P (SEQ ID NO: 8) and *att*P/*att*B (SEQ

ID NO: 9) formed when the *att*P site of pMLP1 is integrated into the (*att*B) site of *M. halophitica.* Table 1 shown below lists isolated pMLP1 sequences and the pMLP1-*M. carbonacea* and pMLP1 - *M. halophitica* *att*B/*att*P and *att*P/*att*B regions as well as their functions based on BLAST homologies.

TABLE 1

GENE PRODUCT OR DNA REGION	SEQ ID NO.	FEATURE	BASE PAIRS
pMLP1 *int*	1	integrase	1394-2572
pMLP1 *xis*	2	excisionase	963-1388
pMLP1 *att*P site	3	*att*P site	2691-2715
*M. carbonacea att*B region	4	tRNA-His	44-119
		*att*B site	95-119
		IR1	124-137
		IR2	142-155
*M. carbonacea att*B/*att*P region	5	tRNA-His	44-119
		*att*B/*att*P site	95-119
		IR1	157-174
		IR2	179-198
*M. carbonacea att*P/*att*B region	6	*att*P/*att*B site	101-125
		IR1	130-143
		IR2	148-161
*M. halophitica att*B region	7	tRNA-His	45-121
		*att*B site	96-121
		IR1	134-145
		IR2	150-161
*M. halophitica att*B/*att*P region	8	*att*B/*att*p site	96-120
		tRNA-His	45-120
		IR1	140-158
		IR2	162-178
*M. halophitica att*P/*att*B region	9	*att*P/*att*B site	101-126
		IR1	139-150
		IR2	155-166

IR = Inverted Repeat

Although the exact sequences of the DNA sites, regions and functionalities are set out in SEQ ID NOS: 1-9, this invention also relates to slight variants of these sequences. Specifically, while the specific sequences are derived from pMLP1, and from *M. carbonacea* and *M. halohitica* transformed with nucleotide sequences derived from pMLP1, the invention encompasses sequences that are homologous or complementary to the sequences as well as sequence-and function-conservative variants to the sequences. Sequence-

conservative variants are those in which a change of one or more nucleotides in a given codon position results in no alteration in the amino acid encoded at that position. Function-conservative variants are those in which a given amino acid residue in a polypeptide has been changed without altering the overall configuration and function of the native polypeptide, including, but not limited to, replacement of an amino acid with one having similar physico-chemical properties (such as, for example, acidic, basic, hydrophobic, and the like). Function-conservative variants also include any polypeptides that have the ability to elicit antibodies specific to a designated polypeptide.

In seeking to protect their invention, the novel sequences have been described in terms of specific sequences as well as those sequences sharing considerable homology to their sequences. Specifically, the invention is intended to include not only the exact polynucleotide sequences as set forth in SEQ ID. NOS: 1-9, but also to include polynucleotide sequences having at least about ninety percent homology to their novel isolated sequences. Preferably, the sequences of the instant invention share at least ninety-five percent homology to the sequences set forth in SEQ ID NOS: 1-9 and most preferably, share at least 98% homology to the sequences set forth in SEQ ID NOS: 1-9 including complete protein coding sequences and complements thereof.

Stringency of conditions employed in hybridizations to establish homology are dependent upon factors such as salt concentration, temperature, the presence of organic solvents, and other parameters. Stringent temperature conditions usually include temperatures in excess of about 30°C, often in excess of about 37°C, typically in excess of about 45°C, preferably in excess of about 55°C, more preferably in excess of about 65°C, and most preferably in excess of about 70°C. Stringent salt conditions will ordinarily be less than about 1000 mM, usually less than about 500 mM, more usually less than about 400 mM, preferably less than about 300 mM, more preferably less than about 200 mM, and most preferably less than about 150 mM. For example, salt concentrations of 100, 50 and 20 mM are used. The combination of the foregoing parameters, however, is more important than the measure of any single parameter. See, e.g., Wetmur *et al.*, J. Mol. Biol. **31**:349 (1968).

VECTORS

The sequences of the invention may be used in any actinomycete into which the vectors of the invention are capable of integrating. For instance, the sequences of the

invention may be incorporated into strains of *Streptomyces*, *Mycobacteria*, *Bacilli*, *Micromonospora* and the like. Strains such as *S. pristinaespiralis* (ATCC 256486), *S. antibioticus* (DSM 40868), *S. bikiniensis* (ATCC 11062), *S. parvulus* (ATCC 12434), *S. glauescens* (ETH 22794), *S. actuosus* (ATCC 25421), *S. coelicolor* (A3(2)), *S. ambofaciens*, *S. lividans*, *S. griseofuscus*, *S. limosus* are particularly useful in fermentation processes. (*See* also, *Smokvina et al.,* Proceedings, 1:403-407).

Vectors that can be used in this invention include microbial plasmids, viruses, bacteriophage, integratable DNA fragments, and other vehicles which may facilitate integration of the nucleic acids into the genome of the host. Plasmids are the most commonly used form of vector but all other forms of vectors which serve an equivalent function and which are, or become, known in the art are suitable for use herein. See, e.g., Pouwels *et al.*, *Cloning Vectors: A Laboratory Manual,* 1985 and Supplements, Elsevier, N.Y., and Rodriguez *et al.* (eds.), *Vectors: A Survey of Molecular Cloning Vectors and Their Uses,* 1988, Buttersworth, Boston, MA.

Expression of nucleic acids utilizing the vectors of this invention can be carried out by conventional methods. Strains of *E. coli* and various actinomycete strains such as *Streptomyces* and *Micromonospora* strains are particularly preferred.

The invention provides a site specific integrase gene and integration site and attachment site. (SEQ ID NOS: 1-3). Use of the *att/int* site-specific integration functions allows for increasing a given gene dosage and for adding heterologous genes that may lead to the formation of new products, such as hybrid antibiotics. This procedure has many advantages over methods involving autonomously replicating plasmids. In particular, *att/int* derived vectors integrate as a single copy per chromosome. Plasmids comprising the site-specific integrating functions allow integration of the gene of choice into the chromosome of actinomycetes. Integrated vectors are extremely stable which allows the gene copies to be maintained without antibiotic selection.

Plasmids comprising the site-specific integrating function of the invention can be used to permanently integrate copies of a heterologous gene of choice into the chromosome of many different hosts. The vectors can transform these hosts at a very high efficiency. Because the vectors do not have actinomycete origins of replication, the plasmids cannot exist as autonomously replicating vectors in actinomycete hosts. The plasmids only exist in their integrated form in these hosts. The integrated form is extremely stable which allows

the gene copies to be maintained without antibiotic selective pressure. The result is highly beneficial in terms of cost, efficiency, and stability of the fermentation process.

Advantageously, the integrative vectors derived from this novel integrase also may comprise a recombinant DNA sequence coding for a desired product, including but by no means limited to, an actinomycete gene. The product can be a peptide, polypeptide or protein of pharmaceutical or agrifoodstuffs importance. One can increase the copy number of the product's sequence per cell, and hence increase the levels of production of a given product. One may also create integrative vectors utilizing the *att/int* genes of the invention to block the biosynthesis of a metabolite, or to produce derivatives of the metabolite.

In addition to using integrating vectors to integrate genes which increase the yield of known products or generate novel products, such as hybrid antibiotics or other novel secondary metabolites, vectors can also be used to integrate antibiotic resistance genes into strains in order to carry out bioconversions with compounds to which the strain is normally sensitive. The resulting transformed hosts and methods of making the antibiotics are within the scope of the present invention.

Prokaryotic expression control sequences typically used include promoters, including those derived from the β-lactamase and lactose promoter systems [Chang *et al., Nature* ***198***:1056 (1977)], the tryptophan (trp) promoter system [Goeddel *et al., Nucleic Acids Res.* ***8***:4057 (1980)], the lambda P_L promoter system [Shimatake *et al., Nature* ***292***:128 (1981)] and the tac promoter [De Boer *et al., Proc. Natl. Acad. Sci.* USA ***292***:128 (1983)]. Numerous expression vectors containing such control sequences are known in the art and available commercially.

Those skilled in the art will readily recognize that the variety of vectors which can be created utilizing the genes of the invention is virtually limitless. The only absolute requirement is that the plasmid comprise an origin of replication which functions in the host cell in which constructions are made, such as *E. coli* or *Bacillus*. No actinomycete origin of replication is required. In fact, in a specific embodiment the plasmid comprising the integrase comprises no actinomycete origin of replication. Other features, such as an antibiotic resistance gene, a multiple cloning site and cos site are useful but not required. A description of the generation and uses of cosmid shuttle vectors can be found in Rao *et al.*, (Methods in Enzymology, 1987, **153**:166198). In short, any plasmid which comprises the integrase is within the scope of this invention.

EXAMPLE II
Construction of *E. coli-Micromonospora* insertion vector pSPRH826b and integration vector pSPRH840

The pSPRH826b insertion vector (Fig. 1) was constructed as follows. A 1.1 kb *Nru*I/*Not*I fragment containing HmR from p16R1 (Garbe *et al., Microbiology*, 1994, **140**:133-138) was treated with T4 polymerase to and ligated to S*sp*I digested, T4 DNA polymerase treated pUC19 to yield pSPRH825. A 787 bp *Pst*I fragment from pRL1058 (oriT region) was treated with T4 polymerase and ligated to N*de*I digested, T4 DNA polymerase treated pSPRH825 to yield pSPRH826b.

The pSPRH840 integrating vector (Fig. 2) was constructed as follows. A 4.0 kb *Kpn*I fragment from cosmid pSPR150 containing the *M. carbonacea* pMLP1 *xis*, *int*, and *att*P region was ligated with *Bam*HI cleaved pBluescriptII KS (Stratagene, LaJolla, CA) to yield pSPRH819. Sequence analysis of the 4.0 kb, *Kpn*I fragment from the cosmid revealed the presence of an integrase gene designated *int*, an excisionase gene designated *xis*, and an integrase attachment site designated *att*P (SEQ. ID NOS: 1-3). A 2.5 kb *Nru*I to *Xho*I fragment from pSPR819 was treated with T4 polymerase, alkaline phosphatase treated and ligated to pCRTopo 2.1 vector (Invitrogen Corp, Carlsbad CA) to yield pSPRH853. A 2.6 kb *Kpn*I to *Pst*I fragment from pSPRH853 was ligated to *Kpn*I and *Pst*I digested pSPR826b (Fig. 1) to yield pSPRH840 (Fig. 2).

EXAMPLE III
Transformation and Integration of pSPRH840 into *M. carbonacea* var *africana* ATCC39149 and *M. halophytica* SCC760

The plasmid pSPRH840 was transformed into *Micromonospora carbonacea* var. *africana* ATCC39149 and *M. halophitica* SCC760 as described in detail as follows.

Micromonospora carbonacea var. *africana* ATCC39149 and *M. halophitica* SCC760 were transformed with pSPRH840 (Fig. 2) by conjugation from *E. coli* S17-1 (Mazodier *et al., Journal of Bacteriology*, 1989, **6**:3583-3585) to *M. carbonacea* and *M. halophitica*. *E. coli* S17-1 containing pSPR840 was grown overnight at 37°C in LB supplemented with 100ug/ml Ampicillin (Amp). The culture was inoculated into LB containing 100ug/ml Amp at an 1:50 ratio and grown with shaking at 37°C to an OD_{600} of 0.4 to 0.5. Cells were harvested by centrifugation and washed three times with fresh LB lacking Amp.

M. carbonacea and *M. halophitica* were grown separately in TSB medium at 30°C with shaking to stationary phase. *E. coli* S17-1 containing pSPRH840 prepared as described above was mixed separately with *M. carbonacea* and *M. halophitica* in a total volume of 100ul plated on AS1 plates using a plastic hockey spreader. Plates were incubated 15hr at 29°C and then overlaid with 50ug/ml naladixic acid and 200ug/ml hygromycin for selection. Transconjugants appearing in 2-3 weeks were picked, homogenized and grown in TSB media with 50ug/ml-naladixic acid and 200ug/ml hygromycin.

Transconjugants appearing in two to three weeks were picked, homogenized, and grown in TSB medium supplemented with 50μ/ml naladizic acid and 200 μ/ml hygromycin. DNA was prepared from *M. carbonacea* transconjugant strains SPH201, SPH202, and SPH207 cleaved with *Kpn*I, separated by gel electrophoresis, a Southern blot prepared, and probed with radiolabled pSPRH819 4.4 kb *Kpn*I insert. Southern hybridization analysis confirmed the presence of a 3.0 kb hybridizing fragment in *M. carbonacea* strains SPH201, SPRH202 and SPRH207 and hybridizing fragments in *M. halophitica* strains transconjugant SPH206, SPH208, and SPH213. Southern analysis of parental *M. carbonacea* showed hybridization to a predicted 4.4 kb *Kpn*I fragment from the replicating pMLP1 phage and 3.5 kb chromosomal fragment. The 3.5 kb hybridizing fragment indicates that pMLP1 is also integrated into the *M. carbonacea* chromosome. *M. carbonacea* strains SPH201, SPRH202 and SPRH207 lacked both freely replicating pMLP1 and integrated pMLP1 hybridizing fragments indicating that pMLP1 and integrated pMLP1 are cured from these strains.

The integration of pSPRH840 into the *M. carbonacea* or *M. halophitica* chromosome forms an *att*B/*att*P left juncture and an *att*P/*att*B right juncture region (Fig. 3, Fig. 4, Fig. 5). These regions were cloned by digesting pSPRH840 integrant strain chromosomal DNA with *Pst*I or *Kpn*I, ligating digested DNA and transforming *E. coli* XL10 (Stratagene, LaJolla, CA). *E. coli* transformants were isolated, plasmid DNA prepared and analyzed by digestion and gel electrophoresis. Plasmids pSPRH873-13 and pSPRH870-1 were identified that contained the *M. carbonacea* *att*B/*att*P and *att*P/*att*B right juncture regions (Fig. 4c, 4d). Plasmids pSPRH871-7 and pSPRH872-10 were identified that contained the *M. halophitica* *att*B/*att*P left and *att*P/*att*B right juncture regions (Fig. 5c, 5d).

The *att*B/*att*P left and *att*P/*att*B right juncture regions formed during pSPRH840 integration in *M. carbonacea* and *M. halophitica* pSPRH840 integrants were sequenced (Fig. 4c, 4d, Fig. 5c, 5d). Analysis confirmed integration of pSPRH840 into the *M.*

carbonacea chromosome, specifically into an *att*B site (bp 95 to bp 119, Fig. 4b) located at the 3' end of a tRNA-His gene (bp 44 to bp 119, Fig. 4b). The *att*P site (bp 101-125, Fig. 4a) and the *att*B site (bp 95-119, Fig. 4b) share perfect homology with each other. During integration an integrative crossover occurs between *att*P located on pSPRH840 and *att*B located on the *M. carbonacea* chromosome. This integrative crossover event creates two regions which contain an *att*B/*att*P site (bp 95-119, Fig. 4c) and an *att*P/*att*B site (bp 101-125, Fig. 4d). The *att*B/*att*P and *att*P/*att*B sites share perfect homology with the *att*P and *att*B sites. The *M. carbonacea* *att*B/*att*P left juncture region is illustrated in Fig. 4c. 5' DNA regions originate from the chromosomal *att*B region (Fig. 4b) and 3' DNA (apparent from bp 137 onward originate from the pSPRH840 *att*P region (Fig. 4a). The *att*B/*att*P left juncture region forms a new functional full length tRNA-His gene (bp 62 to bp 137, Fig. 4c) and an *att*B/*att*P site (bp 95-119, Fig. 4c). The *M. carbonacea* *att*P/*att*B right juncture region is illustrated in Fig. 4d. 5' DNA originates from the *att*P region (Fig. 4a) and 3' DNA (apparent from bp 125 onward) originate from the *att*B region (Fig. 4b). This *att*P/*att*B right juncture region forms an *att*P/*att*B site (bp 101 to bp 125, Fig. 4d) which is a truncated form of tRNA-His containing only the 3' end of the tRNA-His gene.

Analysis also confirmed integration of pSPRH840 into the *M. halophitica* chromosome, specifically into an *att*B site (bp 96. . .120, Fig. 5b) located at the 3' end of a tRNA-His gene (bp 45. . .120, Fig. 5b). The *att*P site (bp 101. . .125, Fig. 5a) and the *att*B site (bp 96. . .120, Fig. 5b) share perfect homology with each other. During integration an integrative crossover occurs between *att*P located on pSPRH840 and *att*B located on the *M. halophitica* chromosome. As is true for the *M. carbonacea* integrants, this integrative crossover event creates two regions which contain an *att*B/*att*P site (bp 96. . .120, Fig. 5c) and an *att*P/*att*B site (bp 101. . . 125, Fig. 5d) As is true for *M. carbonacea,* the *att*B/*att*P and *att*P/*att*B sites share perfect homology with the *att*P and *att*B sites. The *M. halophitica* *att*B/*att*P left juncture region is illustrated in Fig. 5c. 5' DNA regions originate from the chromosomal *att*B region (Fig. 5b) and 3' DNA (apparent from bp 120 onward) originate from the pSPRH840 *att*P region (Fig. 5a). The *att*B/*att*P left juncture region forms a new functional full length tRNA-His (bp 45. . .120, Fig. 5c) and an *att*B/*att*P site (bp 96. . .120, Fig. 5c). The *M. halophitica* *att*P/*att*B right juncture region is illustrated in Fig. 5d. 5' DNA originates from the *att*P region (Fig. 5a) and 3' DNA (apparent from bp 125 onward) originate from the *att*B region (Fig. 5b). This *att*P/*att*B right juncture region forms an

*att*P/*att*B site (bp 101. . .125, Fig. 5d) which is a truncated form of tRNA-His containing only the 3' end of the tRNA-His gene.

PCR primers PDH504 (5' AGGGCAACAAGGGAAGCGTC 3') and PDH505 (5' GGCGGGGGTGTGGCTATTATT 3') were designed to amplify the *att*B region from *M. carbonacea*. PCR amplification of *M. carbonacea* chromosomal DNA yielded a fragment with homology to tRNA-His (bp 45. . .119, Fig. 4b). Contained within this tRNA-His gene, at the 3' end, is the *M. carbonacea att*B site (bp 95. . .119, Fig. 4b) that has perfect homology to the pMLP1 *att*P site (bp 101. . .125, Fig. 4a). PCR primers PDH 502 (5' TTGTTGGTCCGGCCCGCAACG 3') were designed to amplify the *att*B region from *M. halophitica*. PCR amplification of *M. halophitica* chromosomal DNA yielded a fragment with homology to tRNA-His (bp 45. . .120, Fig. 5b). Contained within this tRNA-His gene, at the 3' end, is the *M. halophitica att*B site (bp 96. . .121, Fig. 5b) that has perfect homology to the pMLP1 *att*P site (bp 101. . .125, Fig. 5a).

While the present invention has been described with reference to one or more particular embodiments, those skilled in the art will recognize that many changes may be made thereto without departing from the spirit and scope of the present invention. each of these embodiments and obvious variations thereof is contemplated as falling within the spirit and scope of the claimed invention, which is set forth in the following claims.